Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5072–5076, May 1995
Biochemistry

Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE

KARL R. CLAUSER*, STEVEN C. HALL*, DIANA M. SMITH†, JAMES W. WEBB*‡, LORI E. ANDREWS*, HUU M. TRAN†, LOIS B. EPSTEIN†§¶, AND ALMA L. BURLINGAME*‖**

*Department of Pharmaceutical Chemistry, †Cancer Research Institute, §Department of Pediatrics, and ‖Liver Center, University of California, San Francisco, CA 94143

Communicated by John A. Clements, University of California, San Francisco, CA, January 18, 1995 (received for review June 29, 1994)

ABSTRACT **We report a general mass spectrometric approach for the rapid identification and characterization of proteins isolated by preparative two-dimensional polyacrylamide gel electrophoresis. This method possesses the inherent power to detect and structurally characterize covalent modifications. Absolute sensitivities of matrix-assisted laser desorption ionization and high-energy collision-induced dissociation tandem mass spectrometry are exploited to determine the mass and sequence of subpicomole sample quantities of tryptic peptides. These data permit mass matching and sequence homology searching of computerized peptide mass and protein sequence data bases for known proteins and design of oligonucleotide probes for cloning unknown proteins. We have identified 11 proteins in lysates of human A375 melanoma cells, including: α-enolase, cytokeratin, stathmin, protein disulfide isomerase, tropomyosin, Cu/Zn superoxide dismutase, nucleoside diphosphate kinase A, galaptin, and triosephosphate isomerase. We have characterized several post-translational modifications and chemical modifications that may result from electrophoresis or subsequent sample processing steps. Detection of comigrating and covalently modified proteins illustrates the necessity of peptide sequencing and the advantages of tandem mass spectrometry to reliably and unambiguously establish the identity of each protein. This technology paves the way for studies of cell-type dependent gene expression and studies of large suites of cellular proteins with unprecedented speed and rigor to provide information complementary to the ongoing Human Genome Project.**

Cloning of genes associated with malignancy generates inevitable excitement in biology and medicine. However, subsequent study at the protein level is clearly necessary to understand the processes by which those genes affect vital biological functions, such as the control of gene expression and regulation of cell-signaling pathways. Two-dimensional (2D) PAGE is preferred for simultaneous separation and visualization of proteins present in cell lysates (1, 2). Protein identification and characterization of functionally important primary structural features are the first steps toward gaining insight into the biological roles of specific proteins (3–6). Furthermore, characterization of the 2D-PAGE map from a cell system facilitates both basic research and clinical diagnosis (3, 7).

Traditional partial-sequencing approaches for identifying proteins isolated by 2D PAGE, such as Edman degradation of electroblotted proteins, often meet with limited success due to N-terminal blockage of many eukaryotic proteins. While peptide cleavage, extraction, and HPLC separation may enable subsequent Edman sequencing of internal peptides (8), these procedures alone are not rapid enough to identify the thousands of proteins in a cell lysate in a timely manner. Because of their speed, sensitivity, and ability to deal directly with mixtures, several recently developed mass spectrometry techniques are rapidly becoming the primary methods for identifying proteins isolated by 2D PAGE (4–6, 9, 10). Previously, our laboratories reported the advantages of using liquid secondary ion mass spectrometry, high-energy collision-induced dissociation (CID) tandem mass spectrometry, Edman sequencing, and 2D PAGE to characterize lipocortin I from human melanoma lysates (4). Peptide mass determination and CID sequencing using sample quantities of ≈100 pmol revealed an acetylated N terminus and an unanticipated acrylamide-modified cysteine.

Recently, we have substantially reduced the detection limits of peptide sequencing through incorporation of continuous flow sample introduction and a scanning charge-coupled device array detector onto our tandem mass spectrometer. The resulting chemical noise reduction and rapid recording of single CID spectra (≈10 s) now enable routine peptide sequencing from sample quantities of 100 fmol to 10 pmol (11, 12). Prior to CID, substantial sample is conserved by exploiting the 1–100 fmol sensitivity of matrix-assisted laser desorption ionization (MALDI), rather than liquid secondary ion mass spectrometry for preliminary mass measurement. MALDI enables the generation of peptide molecular-mass maps from fmol levels of unfractionated protein digests or individual HPLC fractions. These peptide-mass fingerprints may be used to search peptide-mass data bases and predict protein identities without resorting to sequencing (5, 9, 13). However, mass-matching algorithms can find only proteins which are already present in a data base. Obviously, primary structure elucidation is essential to identify unknown proteins, search sequence data bases for homologous proteins, characterize covalent modifications, or design oligonucleotide probes for gene cloning. Taking these points into consideration, we report an integrated strategy for identifying and characterizing primary structural features of proteins isolated by 2D PAGE.

MATERIALS AND METHODS

Protein Isolation, Purification, Digestion, and Edman Sequencing. Preparation of human A375 melanoma lysates, isolation of proteins by 2D PAGE (4, 14), protein electroelution of pooled gel spots, tryptic digestion, HPLC separation, and Edman sequencing were performed as described (4). The procedure of Rosenfeld *et al.* (15) was used to produce an in-gel tryptic digest of spot 42. Control digestions without protein substrate were performed so that trypsin autoproteolysis products could be disregarded.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 2D, two-dimensional; CID, collision-induced dissociation; MALDI, matrix-assisted laser desorption ionization.

‡Present address: Department of Chemistry, Illinois State University, Normal, IL 61790.
¶To whom reprint requests should be sent at the † address.
**To whom reprint requests should be sent at the * address.

EXHIBIT D

Mass Spectrometry. Molecular masses (isotopic average) of all tryptic peptides were determined by analyzing 1/50th of each HPLC fraction with a VG TofSpec MALDI mass spectrometer equipped with a nitrogen laser and operated in the linear mode. Peptides were crystallized in matrices consisting of 100 mM 2,4-dihydroxybenzoic acid/50 mM fucose or a saturated solution of α-cyano-4-hydroxycinnamic acid prepared in 0.1% aqueous trifluoroacetic acid (TFA). The mean mass from all spectra recorded for a particular peptide is reported. All MALDI spectra were externally calibrated by using a standard peptide mixture. High-energy positive ion CID mass spectra were acquired on a Kratos Analytical Instruments Concept IIHH four-sector tandem mass spectrometer equipped with a continuous flow, liquid inlet probe and a scanning charge-coupled device array detector (11). Both MS1 and MS2 were operated at 1000 resolution ($M/\Delta m$) to determine monoisotopic masses. HPLC fractions were concentrated to ≈5 μl and diluted to ≈15 μl with a mixture of aqueous 5% (vol/vol) thioglycerol/5% (vol/vol) acetonitrile/0.1% TFA matrix solution. Samples were introduced into the mass spectrometer source at a flow rate of 3 μl/min. Sequences were deduced from CID spectra by using interactive interpretation software developed in our laboratory (16). Quantitation estimates were based on CID and UV response of standards.

Immunoblot Analysis. 2D immunoblot analysis was performed as described (4) with modifications. Gels were loaded with 400–800 μg of protein lysate and transferred at 1.3 mA/cm² for 2 h, and blots were stained with 0.1% Ponceau S/5% (vol/vol) acetic acid. Stain was removed with 100 mM NaOH prior to antibody staining (data not shown).

Data Base Searching. The OWL protein sequence data base (17) was searched by using BLAST (18) with peptide sequences obtained by high-energy CID analysis or Edman degradation. Masses obtained by MALDI were used by MOWSE (internet e-mail server version 5.1; mowse@dl.ac.uk) to search a peptide mass data base constructed from a theoretical trypsin digest of the entire OWL data base (9). Typical parameters employed were a 15% gel-derived protein-mass tolerance, a 2- or 3-Da peptide-mass tolerance, and a partial cleavage scoring factor of 0.4.

RESULTS AND DISCUSSION

Fig. 1 illustrates a representative 2D preparative gel containing human A375 melanoma proteins from whole cell lysates. The spots selected for mass spectrometric analysis were chosen because of their abundance, reproducibility, and relative isolation. Our strategy for the identification and characterization of these proteins is outlined in Fig. 2. Protein spots from several preparative gels were excised, and spots of identical mass pI were pooled before purification from the gel matrix, digestion with trypsin, and peptide separation by HPLC. The yield and number of peptides recovered from each spot were not always proportional to the number of gel plugs excised or the amount and size of protein present, suggesting differential digestion susceptibility, sample handling loss, and occasional protein comigration. For each spot, sub-pmol aliquots of HPLC fractions were analyzed by MALDI. The list of masses obtained from each protein digest served as the focal point for guiding subsequent experiments for identifying and characterizing primary structural features in the protein(s). The number of peptides recovered from digestion of a protein in a 2D PAGE spot varied and depended on protein size and the occasional comigration of other proteins. Experimentally determined masses were used with MOWSE data-base searching to match theoretical peptide masses and attempt to predict protein identities. HPLC fractions containing peptide masses below 2 kDa were subjected to high-energy CID for peptide sequencing. Protein identities were established by using the determined sequences to search the OWL protein sequence data base by using BLAST. Edman degradation was used to sequence peptides with masses greater than ≈2 kDa, since larger peptides ionize less efficiently by liquid secondary ion mass spectrometry.

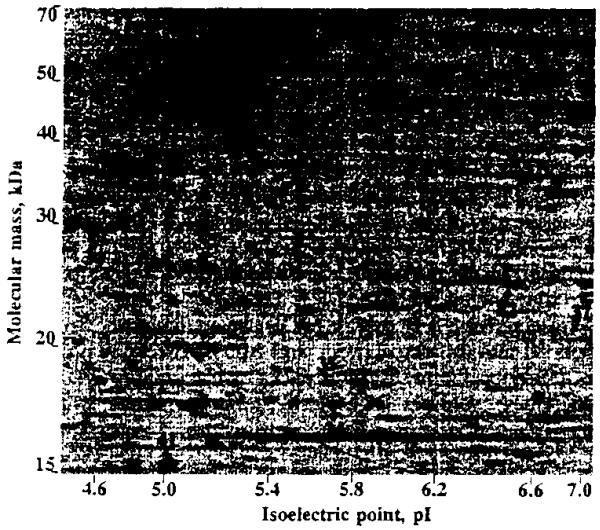

FIG. 1. Coomassie blue G-250-stained 2D preparative gel of human A375 melanoma proteins, numbered as in Tables 1 and 2. One milligram of total protein was loaded. The amount of protein estimated by densitometry in designated spots varied from 0.6 to 2.9 μg.

Tables 1 and 2 summarize peptide mass data, sequences determined or attributed by mass, and data-base search results for all spots studied. Ideally, we seek to attribute every mass to a unique sequence either by sequencing each peptide or matching it by mass to a peptide from the characterized protein.

From the 39 peptide masses from spot 3 listed in Table 1, MOWSE predicted the protein identity as α-enolase. Although most cell types demonstrate α-enolase activity, distribution of the specific enolase isoenzymes in biologically active dimers may be

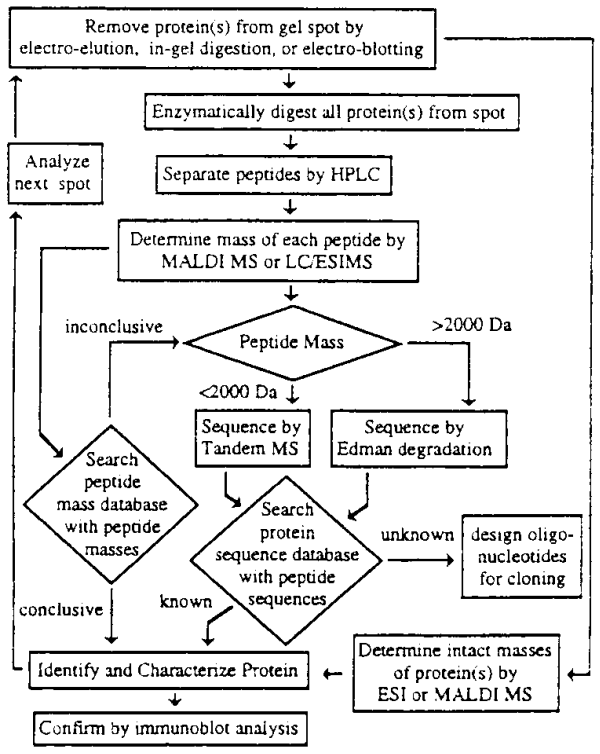

FIG. 2. Strategy for identifying and characterizing proteins separated by 2D PAGE. ESIMS, electrospray ionization mass spectrometry.

tissue specific. Only the α- and γ-homodimers can be detected in serum from patients with malignant ocular melanoma (19). During sequence analysis on selected peptides from spot 3, we also found a cytokeratin sequence LASYLDK that is present in many type I cytoskeletal keratins (mass/pI and peptide masses best matched cytokeratin 15). This indicated that at least two proteins had comigrated. Immunoblot analysis confirmed that α-enolase is present. Furthermore, the presence of a cytokeratin is consistent with the localization of many keratin subtypes to this gel region (20). However, the presence of cytokeratin was not predicted by MOWSE when masses exclusive of those attributable to α-enolase were used.

Protein disulfide isomerase (PDI) and stathmin were found in spots 30 and 24, respectively, and MOWSE readily predicted both. Sequencing of several peptides by high-energy CID established these identities. PDI, which catalyzes formation and interconversion of disulfide bonds in the endoplasmic reticulum, has been implicated in the activation of interferon-inducible genes in chronic myelogenous leukemia cells (21). Immunoblot analysis confirmed the presence of PDI in spot 30 and showed stathmin presence in spot 24 and in two nearby (more acidic) spots. Serine-phosphorylated forms of stathmin have been found in T-lymphocytes (22). The ability to suggest additional posttranslationally modified or genetically variant isoforms of a protein was one motivation for incorporating immunoblot analysis into our strategy.

In spot 33 fibroblast nonmuscle tropomyosin was identified following high-energy CID. Similarly, another form of tropomyosin, cytoskeletal type, was identified in spot 34. The two protein sequences are 76% identical. High-energy CID showed that the masses of the N-terminal peptides from both proteins differed by only 1 Da, due to an internal sequence difference

Table 1. Summary of data obtained for nine human melanoma proteins from eight 2D PAGE gel spots

Spot no.	Protein identified OWL accession no. [Mass/pI] (MOWSE search prediction)	MALDI mass Da	Δ*, Da	Peptide sequence determined by high-energy CID†	MALDI mass, Da	Δ*, Da	Peptide sequence consistent with mass*
3	α-Enolase	1806.5	0.5	(F)AAVPSGASTGIYEALELR(D)‡	1653.5	1.6	(F)EAMRIGAEVYHNLK(N)
	P06733 [45 kDa/6.3]	1406.8	−0.8	(F)GNPTVEVDLFTSK(G)	1542.7	0.9	(F)VVIGMDVAASEFFR(S)
	enoa_human.swiss)	2718.8	2385.9	2042.5 1929.5 1718.5 1643.8§	1424.9	−1.7	(F)YISPDQLADLYK(S)
		2604.9	1503.3	1477.3 1475.3 1436.6 1392.1§	1180.8	−1.6	(K)GVSKAVEHINK(I)
		1385.6	1328.6	1328.5 1268.6 1224.5 1236.1§	1141.6	−2.8	(F)IGAEVYHNLKN)
		1112.7	1095.5	1026.8 984.6 954.7§	905.2	0.2	(F)IEEELGSK(A)
					899.4	−0.7	(K)TIAPALVQK(K)
					705.7	0.8	(K)GVPLYR(H)
	Cytokeratin 15 (type I)	810.4	0.5	(F)LASYLDK(V	1799.1	0.0	(R)VLAEMSEQYLAMAEK(N)
	P19012 [45 kDa/6.3]	1558.7	1044.8¶		1457.6	−1.0	(K)EVASNTEMIQTSK(T)
	not predicted#				1699.2	−2.5	(R)EQYEAMAEK(N)
24	Stathmin	912.6	−0.4	(K)SHEAEVLKQ)	1390.4	0.8	(F)ASGQAFELILSPR(S)
	P16949 [17 kDa/5.7]	888.7	−1.2	( )acASEDIQVK(E)	1350.2	2.7	(K)ESVPEFPLSPPK(K)
	stmn_human.swiss)	1210.5§			1167.1	0.9	(K)AIEENNNFSK(M)
					1040.8	−0.4	(R)KSHEAEVLKQ)
30	Protein disulfide isomerase	939.8	−0.2	(K)TVAYFLQESM)	1583.8	3.0	(F)EATNPPVIQEEKPK(K)
	D16234 [54 kDa/5.8]	867.6	0.9	(K)YQELGEK(D)	1239.1	1.9	(R)DGEEAGAYDGPR(T)
	chumpkal/1.gb_pr)	790.0	2.3	(K)FLDAGHK(L)	1192.1	−0.2	(F)LAPEYEAAATR(L)
		1459.1	1415.3§		998.2	2.0	(K)QAGPASVPLR(T)
33	Tropomyosin,	1247.1	−3.3	(R)IQLVEELDR(A)	1740.8	−3.1	(R)KIQALQQQADEAEDR(A)
	fibroblast non-muscle type	1498.8	−1.7	(M)acAGLNSLEAVKF(K)	1614.7	−1.0	(R)IQALQQQADEAEDR(A)
	P07226 [30 kDa/4.5]	721.0	−1.8	(F)AEEAER(G)	1297.5	−2.1	(R)KLVILEGELER(A)
	tpm_human.swiss)¶	1690.4	1492.7	1427.7 1410.5 1221.2 1192.6§	1272.1	−1.3	(R)EKAEGDVAALNK(R)
		1023.7	893.0	799.3 782.1 696.6 689.8§	1170.0	−2.3	(K)AEGDVAALNRR(I)
					1168.6	−2.8	(K)LVILEGELER(A)
					1014.7	−1.4	(K)AEGDVAALNR(R)
34	Tropomyosin,	1285.5	−0.4	(R)KLVIIEGDLER(T)	1773.1	1.2	(F)KIQVLQQQAEDAEER(A)
	cytoskeletal type	1202.6	0.4	(M)acAGITTIEAVKR(K)	1726.7	−2.2	(R)IQLVEEELDRAQER(L)
	P07224 [29 kDa/4.6]	1155.2	−2.2	(R)LVIIEGDLER(T)	1670.2	−2.7	(K)LVIIEGDLERTEER(A)
	tpm1_human.swiss)	1044.4	−1.0	(M)acAGITTIEAVKR(F)	1472.4	−2.2	(P)IQAEAEVASLNRR(I)
		829.1	−1.9	(R)LEDKLKEL)	1360.9	−0.7	(K)VEEEIKILTDK(L)
		774.2	−1.5	(R)AELAESR(C)	1241.2	−3.2	(F)IQLVEEELDR(A)
		743.1	−1.8	(R)LATALQK(L)	1120.4	−2.0	(R)MELQEDLK(E)
		1958.1	1697.8	1481.0 1465.7 1295.9 1292.3§	911.5	0.2	(R)EMDEQIR(L)
		1255.1	1168.2	1148.2 1073.1 954.3 872.0§	873.7	−1.2	(R)EVEGELRR(A)
35	Cu/Zn superoxide dismutase	822.8	−3.2	(F)TLVVHEKADA)	1023.5	−1.6	(K)HGGPKDEERDH)
	P00441 [18 kDa/5.7]	668.6	−1.7	(K)VWGSIK(G)	777.7	−2.1	(K)crESNGPVK(V)
	not predicted#	2210.4	1972.6	1498.2 1195.8 1127.6 1101.2 1039.2§	962.8	0.7	(K)TGNAGSR(L)
36	Nucleoside diphosphate kinase A	1344.5	−1.3	(F)TFIAIKPDGVQR(G)	1193.1	−0.3	(K)crERPFFAGLVK(Y)
	P15531 [18 kDa/5.8]	1149.2	−1.2	(F)RDFIAGLVK(Y)			
	nmka_human.swiss)#	1066.1	−1.3	(F)GDFCamIQVGR(N)			
		1800.7	825.6	659.4§			
41	Galaptin	1661.2	−1.5	(F)ENAHGDANTIVCamNSK(D)	1585.2	0.6	(K)crDSNNLCamLHFNPR(F)
	P09382 [16 kDa/5.1]	1502.2	9.6	(F)DSNNLCamLHFNPR(F)	1075.6	−1.7	(K)DGGAWGTEQR(E)
	not predicted#	877.1	−0.8	(F)SFVLNLGK(D)	969.0	−0.1	(K)LPDGYEFK(F)
		2400.1	2210.1	2067.0 1993.1 1950.1 1940.4 1849.2§	909.5	−0.5	(K)crIKCamVAFD( )
		1822.0	1802.4	1781.1 1680.1 1525.3 1516.4§			

*The difference between the measured mass and calculated mass (average isotopic).
†Abbreviations: ( ), residues before/after peptide; Cam, acrylamide-modified Cys; Camo, Cam (oxidized); acX, acetylated N terminus; crX, carbamylated N terminus; Mso, Met sulfoxide.
‡Sequenced by Edman degradation.
§Masses neither identified nor attributed.
¶Masses ambiguous: α-enolase or cytokeratin.
#Inconclusive search result.

Table 2. Summary of data obtained for triosephosphate isomerase from 2D PAGE gel spots 37 and 42

Spot no.	Protein identified OWL accession no. (Mass/pI) (MOWSE search prediction)	MALDI mass, Da	Δ*, Da	Peptide sequence determined by high-energy CID†	MALDI mass Da	Δ*, Da	Peptide sequence consistent with mass†
37	Triosephosphate isomerase P00938 [25 kDa/6.4] (tpis_human.swiss)	1647.7	3.0	(R)crD-CamGATWVVLGHSEF R)	1026.6	−4.8	(R)ELAQGTEVDGFLVGGASLKPEFVDIINAK(Q)
		1620.9	−2.0	(F)" TNGAFTGEISPGMIKDV	2208.2	0.7	(K)VPADTEVVCamAPPTAYIDFAR(Q)
		1615.6	−0.2	(F)5 HVFGESDELIGQK(V)	1836.8	−2.8	(R)VAHALAEGLGVIACamIGEK(L)
		1599.2	−2.6	(K)I CamGATWVVLGHSER(F)	1820.6	−2.0	(R)VAHALAEGLGVIACamIGEK(L)
		1595..	−8.8	(F)VVLAYEPVWAIGTGK(T)	1638.9	0.0	(K)VTNGAFTGEISPGMsoIK(V)‡
		1459.0	−0.6	(R)HVFGESDELIGQK(V)	1607.7	1.6	(R)LDPKIAVAAQNCamYK(V)
		1408.7	−6.9	(K)GSLGELIGTLNAAK(V)	159...	0.6	(K)VVFEQTKVIADNVK(D)
		1380.9	−3.7	(R)crHYGGSVTGATCamK(E)	1503.8	1.2	(R)crHVFGESDELIGQK(V)
		1236.2	1.8	(K)ICamGATWVVLGH(S)	1468.7	1.1	(R)TATPQQAQEVHEK(L)
		1196.1	0.7	(K)GIAVAAQNCamYK(V)	1341.5	−0.3	(R)HYGGSVTGATCamK(E)
		1168.7	0.3	(K)IAVAAQNCamYK(V)	1278..	2.7	(K)VIADNVEDWSK(V)
		956.4	1.3	(K)FFVGGNWK(M)	1234.4	−0.9	(K)SNVSDAVAQSTR(I)
		892.5	−1.5	(K)cVVFEQTK(V)	1151.2	−1.2	(K)IAVAAQNCamYK(V)
		758.9	0.0	(K)VIADNVK(D)	1123.3	−3.2	(R)crKFFVGGNWK(M)
		1747.1	1643.5	1481.4 1469.3 1109.4 826.5§	1081.9	−1.4	(R)KFFVGGNWK(M)
42	Triosepeciphate isomerase P00938 [25 kDa/6.4] (tpis_human.swiss)	1760.1	2..	(K)ICamGATWVVLGHSERR H)	2035.3	3.9	(R)ELASQPDVDGFLVGGASLKPEFVDIINAK(Q)‡
		1602.6	0.8	(K)ICamGATWVVLGHSER(F)	2342.9	6.2	(K)VAHALAEGLGVIACamIGEKLDER(E)‡
		1407.7	−7.9	(K)QSLGELIGTLNAAK(V)	2203.9	−3.6	(K)VPADTEVVCamAPPTAYIDFAR(Q)‡
		1340.1	−1.5	(R)HYGGSVTGATCamK(E)	1826.4	3.2	(K)VAHALAEGLGVIACamIGEK(L)‡
		1235.7	0.4	(K)SNVSDAVAQSTR(I)	1615.6	−0.2	(R)RHVFGESDELIGQK(V)
		953.1	−2.0	(K)IFVGGNWK(M"	1604.7	0.8	(K)VVLAYEPVWAIGTGK(T)‡
		2210.7	2089.8	2034.2 1823.1 1594.4 1400.8§	1543.4	−0.4	(R)KQSLGELIGTLNAAK(V)
		1063.0	1060.4§		1467.1	−0.5	(K)TATPQQAQEVHEK(L)
					1457.9	−1.7	(R)HVFGESDELIGQK(V)
					1151.5	−0.9	(K)IAVAAQNCamYK(V)
					1082.6	−0.7	(R)KFFVGGNWK(M)
					757.8	−1.1	(K)VIADNVE(D)
					747.6	−0.8	(F)EAGITEK(V)

*The difference between the measured mass and calculated mass (average isotopic).

†Abbreviations: ( ), residues before/after peptide; Cam, acrylamide-modified Cys; Camo Cam (oxidized), acX, acetylated N terminus; crX, carbamylated N terminus; Mso, Met sulfoxide.

‡Sequenced by Edman degradation.

§Masses neither identified nor attributed.

¶Masses ambiguous α-enolase or cytokeratin.

‖Inconclusive search result.

of ITTI (spot 34) vs. LNSL (spot 33). The presence of N-terminal acetylation is evident in Fig. 3 from the 42-Da increase in the mass of all N-terminally derived *a*- and *b*-type ions. Leucine at residues 3 and 6 is evident in Fig. 3 from the mass values of the *w*6 (656.3 Da) and *w*9 (970.4 Da) ions resulting from side-chain fragmentation. Isoleucine was likewise assigned at those same positions in the corresponding peptide from spot 34 (data not shown).

Differentiation and unambiguous structure assignment of these similar peptides illustrates the power of high-energy CID in protein primary structure determination. Methodology and fragment-ion nomenclature for high-energy CID are more thoroughly described elsewhere (12, 23). Although their biological function in nonmuscle cells is not clear, the tropomyosins are multiple isoforms which complex with microfilaments (24). While spot 24 was readily predicted as cytoskeletal tropomyosin by MOWSE, fibroblast nonmuscle tropomyosin was the only isoform among several ambiguous possibilities predicted for spot 33. Immunoblot analysis confirmed that tropomyosin is present in spots 33 and 34 and two more basic spots nearby.

Nucleoside diphosphate kinase A (NDPK-A) was identified in spot 36 following high-energy CID. NDPK catalyzes the phosphorylation of nucleoside 5′ diphosphates. NDPK-A is the product of the *nm23* gene, which, when overexpressed, decreases tumorigenesis of melanoma cell lines (25). MOWSE results for this spot were inconclusive, due in part to chemical modification of two peptides.

Cu/Zn superoxide dismutase (Cu/Zn SOD) and galaptin were identified in spots 35 and 41, respectively, following high-energy CID. Immunoblot analysis confirmed the presence of both Cu/Zn SOD and galaptin. Cu/Zn SOD catalyzes the conversion of toxic superoxide to hydrogen peroxide. Galaptin, a lectin present in many tissues, has been found in several skin tumor types including melanoma, and its reduced expression level has been suggested as a means of diagnosing malignancy (26). MOWSE was unable to predict the identity of both of these proteins from the obtained MALDI data. Only 15–25% of the MALDI masses obtained for spots 35 and 41 were consistent with unmodified tryptic peptides. Our studies show that for any given MOWSE search, numerous proteins present in the data base appear to be randomly capable of matching ≈30% of the masses in a given list. Below this level we have little confidence in MOWSE predictions and find sequence determination necessary for protein identification.

Triosephosphate isomerase (TPI), present in both spot 37 and spot 42, was readily predicted by MOWSE (Table 2). The success of MOWSE is noteworthy despite the fact that several of the MALDI masses are quite inaccurate (an error of more than ±3.5 Da, attributed to an inconsistently performing laser power supply). Subsequent high-energy CID and Edman degradation elucidated the sequences of several covalently modified peptides and established the identity of TPI in both spots. Peptides that represented a nonspecific trypsin cleavage between His-95 and Ser-96 and chemical modifications, including acrylamide-modified cysteines, oxidation, and carbamylation, were found. The chemical modification of cysteine arose from protein interaction with the gel matrix. Carbamylation is likely a consequence of using 2 M urea buffer during enzymatic digestion of spot 37. In-gel digestion with 100 mM ammonium bicarbonate was done for spot 42 rather than electroelution, thus eliminating the need for urea.

TPI, which catalyzes the interconversion of dihydroxy acetone phosphate and glyceraldehyde 3-phosphate in glycolysis, gluconeogenesis, fatty acid synthesis, and the pentose shunt, is

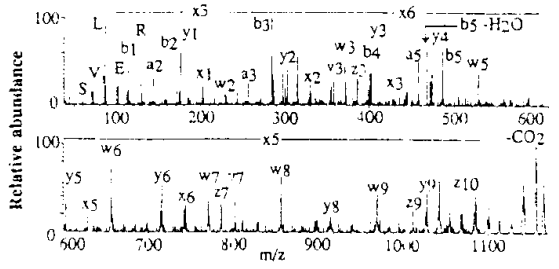

FIG. 3. High-energy CID spectrum of acetylated N-terminal tryptic peptide of nonmuscle tropomyosin (1–10 pmol) from spot 32, MH$^+$ = 1199.7 (monoisotopic mass). Sequence: CH_3CO-AGLNSLEAVKR. Peptide backbone cleavage ions associated with charge retention at the C terminus are denoted by *x*, *y*, and *z*; and at the *N* terminus by *a* and *b*. Side-chain fragment ions are denoted by *v* and *w*.

the product of a single gene. However multiple electrophoretic forms have been observed, and some result from deamidations at Asn-71 and Asn-15 or oxidation of Cys-126 (27). By high-energy CID, an Asn-71 containing peptide in spot 37 (mass 1620.9 in Table 2) was not deamidated. The corresponding peptide was not found in spot 42. All five cysteines in TPI were shown by high-energy CID, Edman degradation, or mass matching to be consistent with acrylamide modification for spots 37 and 42. Throughout this work we have found cysteines exclusively in the acrylamide-modified form, despite inclusion of the antioxidant sodium thioglycolate during electrophoresis.

We have not currently pursued the strategy outlined in Fig. 2 which uses mass spectrometry to determine accurate masses on intact proteins isolated by 2D PAGE. For proteins in solution, mass determinations accurate to 0.01% (±2 Da on a 20-kDa protein) may be obtainable by MALDI or electrospray ionization mass spectrometry (28). Accurate intact masses could suggest the existence and possible identity of posttranslational modifications or confirm an expected protein sequence. However, mass determination on intact proteins isolated by 2D PAGE is confounded by problems associated with removal of stained protein from the gel matrix in a format amenable to mass spectrometry. Initial attempts at MALDI from proteins and peptides blotted onto polyvinylidene difluoride and nylon membranes are encouraging (29–31).

Since our strategy is based on obtaining sequence information at fmol to low pmol levels, discovery of previously unknown peptide sequences will enable design of oligonucleotide probes for gene cloning. Furthermore, the growth of sequence data bases fueled by the Human Genome Project will increasingly reduce the chances of discovering unknown proteins. However, the discriminating power of a peptide mass-matching strategy employing ±2–3 Da MALDI mass data will progressively decline with data-base growth, thus producing increasing numbers of inconclusive results, necessitating sequencing. Enhanced mass accuracy obtained by incorporating electrospray ionization mass spectrometry or improvements in MALDI technology would slow the anticipated decline in discriminating power of mass-matching strategies by reducing the effective portion of the data base considered in a search. Hence, when protein recognition is sought without need for primary structure characterization, the identities of many known proteins may be rapidly predicted by combining mass matching with direct MALDI analysis of unfractionated digests (5, 6, 9, 10). Unfortunately, the ability to directly perform substantial sequence analysis is severely compromised. Of the 11 proteins we characterized from 10 gel spots, 6 were readily predicted by mass matching with MOWSE, and 9 contained unmatched covalently modified peptides. Consequently our results illustrate the necessity of peptide sequencing and the advantages of tandem mass spectrometry to rapidly and unambiguously identify proteins isolated by 2D PAGE.

We thank F. C. Wells for technical expertise, and Fisons Instruments for loan of a VG TofSpec. We are grateful to Dr. S. Hanash for stathmin monoclonal antibody and to Drs. L. Pohl, L. Oberley, J. Zieske, and H. Allen and Chemicon International for antisera. This work was supported by National Institutes of Health Grants DK26743, RR03611 and RR04705 to A.L.B. and Grants CA27903 and CA44446 to L.B.E. National Science Foundation Grant DIR-8700766 to A.L.B.; University of California Tobacco Research and Disease Related Program Grant 2RT0689; and a University of California Systemwide Biotechnology Grant.

1. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
2. Smith, D. M., Tran, H. M. & Epstein, L. B. (1995) in *Cytokines: A Practical Approach*, ed. Balkwill, F. (Oxford Univ. Press, London), in press.
3. Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., *et al.* (1992) *Electrophoresis* **14**, 1091–1198.
4. Hale, S. C., Smith, D. M., Martinez, F. F., Seo, V. W., Tran, H. M., Epstein, L. B. & Burlingame, A. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1427–1431.
5. Hunzel, W. J., Bileci, T. M., Stults, J. T., Wong, S. C., Grimley, C. & Watanabe, C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.
6. Rasmussen, H. H., Moritz, E., Mann, M., Roepstorff, P. & Celis, J. E. (1994) *Electrophoresis* **15**, 406–416.
7. Tessel, A., Paquet, N., Huber, G., Frutiger, S., Tissot, J., Hughes, G. & Hochstrasser, D. (1992) *J. Chromatogr.* **582**, 87–92.
8. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
9. Hoppe, D. J., Højrup, P. & Bleasby, A. J. (1993) *Curr. Biol.* **3**, 327–332.
10. Jones, H., Whitehead, R. H., Reid, G. E., Moritz, R. L., Ward, L. D. & Simpson, R. J. (1994) *Electrophoresis* **15**, 391–405.
11. Wilm, F. C., Hall, S. C., Medzihradszky, K. F., Yu, Z., Burlingame, A. L., Evans, S., Hoffman, A. D., Buchanan, R. & Glover, S. (1993) *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics* (Am. Soc. Mass Spectrom., San Francisco), pp. 927a–927b.
12. Medzihradszky, K. F. & Burlingame, A. L. (1994) *Methods Companion Methods. Enzymol.* **6**, 284–303.
13. Mann, M., Højrup, P. & Roepstorff, P. (1993) *Biol. Mass Spectrom.* **22**, 338–345.
14. Patton, W. F., Pluskal, M. G., Skea, W. M., Buecker, J. L., Lopez, M. F., Zimmerman, R., Belanger, L. M. & Hutch, P. D. (1990) *BioTechniques* **8**, 518–527.
15. Rosenthal, G. J., Capodieci, J., Guillemot, J. C. & Ferrara, P. (1992) *Anal. Biochem.* **203**, 173–179.
16. Jones, W. M., Falick, A. M., Burlingame, A. L. & Gibson, B. W. (1992) *J. Am. Soc. Mass Spectrom.* **3**, 326–336.
17. Bleasby, A. J. & Wootton, J. C. (1990) *Protein Eng.* **3**, 153–159.
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
19. Saone, B. S., Hungerford, J., Vaghela, B. & Sheraidah, G. A. (1990) *Int. J. Ophthalmol.* **74**, 427–430.
20. Sun, T. T., Tseng, S. C., Huang, A. J., Cooper, D., Schermer, A., Lynch, M. H., Weiss, R. & Eichner, R. (1985) *Ann. N.Y. Acad. Sci.* **455**, 307–329.
21. Johnson, E., Henzel, W. & Deisseroth, A. (1992) *J. Biol. Chem.* **267**, 13412–13421.
22. Strahler, J. R., Hanash, S. M., Lamb, B. J., Rogers, K. P., Underhill, J. A., Meilhem, R. F., Kelm, D. R., Zhu, X., Kuick, R. D., Fox, D. A. & Hanash, S. M. (1992) *J. Immunol.* **149**, 1191–1198.
23. Biemann, K. (1990) *Methods Enzymol.* **193**, 455–479.
24. Goes-Miller, J. & Helfman, D. (1991) *BioEssays* **13**, 429–437.
25. Leone, A., Flatow, U., King, C. R., Sandeen, M. A., Margulies, I. M., Liotta, L. A. & Steeg, P. S. (1991) *Cell* **65**, 25–35.
26. Santa-Lucia, P., Wilson, B. D. & Allen, H. J. (1991) *J. Dermatol. Surg. Oncol.* **17**, 653–655.
27. Gracy, R. W., Yuksel, K. U., Chapman, M. L. & Dimitrijevich, S. D. (1990) *Prog. Clin. Biol. Res.* **344**, 787–817.
28. Chait, B. T. & Kent, S. B. (1992) *Science* **257**, 1885–1894.
29. Strupat, K., Karas, M. & Hillenkamp, F. (1994) *Anal. Chem.* **66**, 464–470.
30. Westling, M. M. & Henselau, C. (1994) *Anal. Chem.* **66**, 471–477.
31. Zaluzec, E. J., Gage, D. A., Allison, J. & Watson, J. T. (1994) *J. Am. Soc. Mass Spectrom.* **5**, 230–237.